

Weekly Report

1 Done

1.1 Paper Revision

After discussing with Chris, I re-write some paragraphs.

1.2 Learning SGD

Stochastic gradient descent (SGD) is an iterative method for optimizing an objective function with suitable smoothness properties.

At the beginning, SGD randomly shuffles examples in the training set. Following the random order, SGD updates parameters as following:

$$\theta^{(t)} \leftarrow \theta^{(t-1)} - \epsilon_t \frac{1}{B} \sum_{t'=Bt+1}^{B(t+1)} \frac{\partial L(z_{t'}, \theta)}{\partial \theta}.$$

As batch size B equals 1, this process is standard gradient descent, but not stochastic. As $1 < B < \text{the training set size}$, this process is called minibatch gradient descent. Batch size determines how many examples you look at before making a weight update. The lower it is, the noisier the training signal is going to be, the higher it is, the longer it will take to compute the gradient for each step. Each time we take a sample and update our weights it is called a mini-batch. Each time we run through the entire dataset, it's called an epoch.

SGD can be applied to the datasets with dynamically increased records.

1.3 Mooc Verification

1.4 Book Verification